



OPEN

Phylogenetic structure of *Salmonella* Enteritidis provides context for a foodborne outbreak in Peru

Willi Quino¹, Junior Caro-Castro¹, Orson Mestanza¹, Carmen V. Hurtado¹, Maria L. Zamudio¹ & Ronnie G. Gavilan^{1,2}✉

Salmonella Enteritidis, an important foodborne zoonosis, has a dramatically increased number of cases around the world. To explore the phylogenetic structure of Peruvian *Salmonella* Enteritidis strains and their relationship with an outbreak occurred in 2018, we analyzed a comprehensive strains of *S. Enteritidis* received by the National Institute of Health during the period 2000–2018. A total of 180 strains were characterized by microbiological procedures, serotyping and whole genome sequencing. Based on genome sequences annotated, virulence factors and accessory genes were identified. Phylogenetic and population structure analysis were also analyzed based on SNPs. The phylogenetic analysis grouped the genomes into two well-supported clades that were consistent with population structure analysis. The clinical and food strains corresponding to the outbreak were included in the same cluster, which presented the *sdhA* gene, related to the increase of the virulence of this pathogen. The phylogenetic relationship of Peruvian *S. Enteritidis* suggests the presence of four *S. enteritidis* population with high epidemiological importance.

Salmonella spp. are one of the major causes of morbidity and mortality in children in many developing countries worldwide. The impact on global human health of non-typhoidal salmonellosis (NTS) is high, with an estimated of 93.8 million cases. 80.3 million of cases were related to food transmission¹. In Latin America, Asia and Africa, the reported incidence of *Salmonella* spp. is 200 to 500 cases per 100,000 inhabitants annually. Since the transmission of *Salmonella* spp. from person to person is uncommon, food is considered the main source of human exposure, with an estimated 95% of infections related to animal source food².

S. enterica subsp. *enterica* serovar Enteritidis (*S. Enteritidis*) is a serovar that colonizes the reproductive tract of chickens, including other birds and wild animal, usually asymptotically³. In addition, the bacteria use bird eggs as the main vehicle of infection, which can be contaminated internally and on the outer shell surface. Internal contamination is usually due to the result of bacterial penetration through the egg shell via microscopic cracks. On the other hand, external contamination occurs before oviposition, caused by infection of the bird's reproductive organs⁴. In addition, there are other factors that increase *S. Enteritidis* infection among humans, such as cross contamination of raw and prepared foods⁵, as well as the existence of asymptomatic infected food handlers⁶. For all these reasons, an emerging foodborne zoonosis caused by *S. Enteritidis* has increased dramatically since the 1980s, when *S. Enteritidis* replaced *S. Typhimurium* as the primary cause of salmonellosis worldwide⁷. In fact, recent studies in Latin America from 2005 to 2009 reported a higher *S. Enteritidis* isolation rate (51%) compared to 21% *S. Typhimurium*⁸.

Whole genome sequencing (WGS) is currently being widely used for the surveillance of foodborne pathogens. Prospectively, the typing of *S. Enteritidis* strains could identify possible outbreaks in real time. Phylogenetic methods that explore the relationship between microbial genomes have been used for the study of emergence, geographic spread and transmission of infections⁹, as shown by various investigations related to this pathogen^{3,10,11}. Therefore, in this study we explored the phylogenetic structure of Peruvian *S. Enteritidis* strains from the last 20 years and its relationship with an outbreak occurred in Callao, Peru, in 2018.

¹Centro Nacional de Salud Pública, Instituto Nacional de Salud, Lima, Peru. ²Escuela Profesional de Medicina Humana, Universidad Privada San Juan Bautista, Lima, Peru. ✉email: rgavilan@ins.gob.pe

Results

Identification of *Salmonella enteritidis*. In the study, all 180 strains, including strains obtained from the outbreak, were identified as *S. Enteritidis* by conventional microbiological procedures, whereby all the strains were defined by the formula 1,9,12: g, m:— (Supplementary material 1). The geographical area where each strain was isolated can be seen in Fig. 1A, colored by the populations obtained with hierBAPS.

Analysis of *Salmonella Enteritidis* genomes. Genome sequencing analysis was performed with all 180 strains, obtaining an average of 3.3 million reads, 77 contigs and a G + C percentage of 52.0%. The average size of the genomes was 4.8 Mb. MLST analysis determined that all sequenced strains of *S. Enteritidis* corresponded to the sequence type 11 (ST11), which is the most common ST of *S. Enteritidis* (http://enterobase.warwick.ac.uk/species/senterica/search_strains).

The alignment of the 180 genome sequences gave us a core genome of 65.7%, which presented 6208 SNPs within locally colinear blocks of more than 3.2 Mb. The ratio of recombination and mutation was $R/\theta = 0.07$, whereas the ratio of effects of recombination and mutation was $r/m = 0.26$, having low recombination events in these bacteria.

The Maximum Likelihood tree (Fig. 1B) revealed two well-defined clades: clade A, encompassing the majority of strains (143), including those belonging to the outbreak, and clade B, with 37 strains. The phylogeny was divided into four populations using hierBAPS, all with 100% of bootstrap. Clade A is composed only by population 1 (green), whereas clade B comprises population 2 (orange), 3 (blue) and 4 (red).

Population 1 (green) contains strains from clinical, environmental and animal sources from 2001 to 2013. It is noteworthy that only strains isolated from 2001 to 2002 belong to population 1, corresponding to 48.6%. Population 2 (orange) comprises two environmental strains isolated in 2017, whereas population 3 (blue) contains clinical (94.3% of the total number of strains) and environmental strains, obtained from 2000 to 2017. Population 4 (red), located in the branch most internal one, comprises clinical, food and environmental strains obtained from 2008 to 2018. The recent outbreak strains are located inside population 4.

Analysis of virulence factors and accessory genome. Few genes gave relevant information that allowed us to find differences between clades. *pefB* and *spvB* genes were found in almost all Peruvian strains (98.9% and 84.4%, respectively), and did not show any evident pattern. A similar situation was found with *ssaI* gene, which was detected in half of strains evaluated (50.6%). However, the *shdA* gene was absent in the majority of studied strains (91.7%), with the particular exception of those belonging to the 2018 outbreak, which presented it within its genome. Regarding to phage regions ϕ SE20, they were found in a big number of Peruvian strains (80.6%). The presence or absence of these genes can be visualized attached to each strain in Fig. 1B.

Discussion

Salmonella Enteritidis is one of the main causes of food-related illness worldwide, facilitated by its ability to colonize and infect a wide range of hosts, especially through eggs and products made with them²². In Peru, there are few studies and records related to the study of the epidemiology of this serovar, and little information related to phylogenetic and molecular studies^{23,24}. Therefore, this study offers a first look at the phylogenetic structure of Peruvian *S. Enteritidis* strains received by INS from different regions of Peru as part of national surveillance.

In this work, genomes of *S. Enteritidis* isolated in Peru from diverse geographical origins and diverse sources of isolation in a period of almost 20 years were analyzed, which presented a very similar genetic profile, independently of the source from which they were obtained or their host. It is known that *S. Enteritidis* is a highly clonal serovar within the genus *Salmonella*²⁵. In our study, all strains belonged to the same ST and showed low rate of recombination, which reflects the high degree of clonality of this circulating serovar through Peru. The maximum likelihood analysis of Peruvian strains shows two clearly differentiated clades, although no specific grouping depends on the source of isolation: clade A, whose branch is close to the outgroup *S. Gallinarum*, and clade B, whose branch is far from the reference used as outgroup. This result is a pattern that is often repeated in many phylogenetic works related to this serovar^{11,26}; consequently, it could be an evidence that two lineages of *S. Enteritidis* are currently circulating in Peru.

A relevant fact is that strains from different sources can be phylogenetically related. This is especially relevant in the analysis of the most internal population obtained with the hierBAPS (population 4, red color). The isolated strains from mayonnaise are phylogenetically related to the strains obtained from the patients of the outbreak, which incriminates mayonnaise as a transport vehicle of the disease. In addition, inside the genetic information of strains from the outbreak, we emphasize the presence of *shdA* gene. This gene encodes an intestinal colonization adhesin, homologous to various adhesins found in other pathogens such as diffuse adhesion *E. coli* (DAEC), *S. Typhimurium* and *Shigella flexneri*. Also, it is commonly associated with *Salmonella* serovars isolated from poultry, being involved in the expansion of *S. enterica* host range within warm-blooded vertebrates²⁷. Besides, it is the only known determinant for persistence bacterial in the cecum, allowing an efficient and prolonged excretion of it in the feces, which contributes to a greater virulence of this pathogen^{28,29}.

A detail in the outermost population (population 1, green color) is remarked: the presence of isolated strains from 2001–2002, which make up half of strains from this group. These strains are not found in the other populations. In addition, other members of this population after the indicated years suggest that the circulation of strains of this lineage is still occurring in a small number. This could explain the difficulty to trace them the year by year. Inside population 2 (orange color), there are only two strains of environmental source. This could indicate a high rate of variation outside of an animal host.

Another interesting result was the presence of the phage ϕ SE20 regions, which was found in 80.4% of the strains, and located in the four populations obtained in the analysis. Shah et al.³⁰ define it as a prophage that

transports genes involved in the pathogenicity of *Salmonella*, which is suspected to be related to the emergence of *S. Enteritidis* worldwide. Feasey et al.³¹ found the presence of these regions within what they called the “global epidemic clade”. Subsequently, in a study made in Brazil, the presence of these regions was reported in all strains after 1994, year in which the increase of serovar *Enteritidis* strains began within this country, concluding that the emergence of the infection by this serovar could be related to that phage³. Due to previous reports, and the findings of the present study, the emergence of *S. Enteritidis* in Peru could be related to the presence of strains that include phage regions ϕ SE20 within its genome. To have more support for this hypothesis, it is required to obtain samples prior to 1995, and to verify the majority absence of these regions in those strains.

In a complementary way, it is observed that the distribution of the strains by geographical region of Peru does not follow a pattern, because strains from different populations are distributed indistinctly between the different regions. Therefore, it can be deduced that both lineages would be causing infection throughout the territory. This same distribution that does not follow any clear distribution was also previously observed in Brazil and Chile^{3,11}.

The study presents the limitation of not having isolated strains prior to the year 2000 and from the 25 regions of Peru, which would have helped significantly to represent the overall picture of the country. Also, it could help to reinforce the hypothesis of Campioni et al.³, where the majority absence of genes belonging to phage ϕ SE20 was highlighted, a stage called pre-pandemic by these authors.

In conclusion, the phylogenetic relationship of Peruvian strains from 2000 to 2018 suggests the presence of two well-defined clades and, therefore, two different lineages of *S. Enteritidis* causing disease in humans. The presence of four well-defined populations with particular characteristics suggest discrete groups of clinical strains that are being transmitted in the different geographical areas without showing any clear pattern of geographical distribution is remarked. Finally, the outbreak occurred in 2018 in Callao was due to a highly virulent strain, evidenced by the genomic presence of the *shdA* gene, whose frequency rate in strains from the period of time evaluated is low.

Methods

Sampling and population. In this study, we included a total of 180 strains received by the National Reference Laboratory of Enteropathogens of the National Health Institute (in Spanish, Instituto Nacional de Salud, INS) from different parts of Peru during the years 2000–2018, which were previously characterized as *Salmonella* spp.

Nine strains were isolated in order to study a possible outbreak occurred in Callao, Peru. These nine strains included five isolated from patients with signs and symptoms of acute diarrheal disease after consuming at a restaurant in Callao, Peru, occurred in 2018, and three strains from the possible source of contamination, mayonnaise prepared in the same place.

Ethical statement. The use of human biological material was approved by the Institutional Research Ethics Committee/INS with register No 20516–2018. The patient information was completely anonymized in this study.

Culturing and biochemical confirmation. All 180 strains were recovered using a selective media. First, all the strains were inoculated in Trypticase Soy Broth (Merck, Germany) at 37 °C from six to eight hours. Subsequently, they were grown on SS Agar plates (Salmonella Shigella Agar; Merck, Germany) and incubated at 37 °C from 18 to 24 h. The genus *Salmonella* was confirmed using conventional biochemical tests. These strains were analyzed with O and H agglutination antisera (Bio Merieux, France) following the Kauffmann-White scheme for the identification of serovars¹².

Preparation of libraries and sequencing. All 180 strains identified by serotyping as *Salmonella* *Enteritidis* were selected for sequencing. The DNA extraction of those strains was carried out using the DNeasy Blood & Tissue kit (Qiagen, Germany) following the manufacturer’s instructions. The genomic DNA samples previously obtained were quantified by fluorometry (Qubit 3.0 Invitrogen, Malaysia). The sequencing libraries was performed using the Nextera XT Library Preparation Kit (Illumina, USA), and genomic sequencing using MiSeq (Illumina, USA).

Analysis of genomic data. The quality of each obtained sequence was evaluated through the software FastQC v0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Adapters and low-quality bases were removed using the program Trimmomatic v0.38¹³. The sequences were assembled de novo using the pipeline A5-miseq v20160825¹⁴. The identification of genus and detection of possible contaminated contigs were made with the program Kraken v1.0¹⁵ using Minikraken database. Inclusion criteria include genomes with had less than 300 contigs and 92% of them belonging to *Salmonella* spp. The genotype of each sequence was identified using the software Multilocus Sequence Typing (MLST) v2.10.

For the phylogenetic study, the genome of *S. Gallinarum* strain 287/91 (accession number AM933173.1) stored in the GenBank database (<http://www.ncbi.nlm.nih.gov>) was used as outgroup. Subsequently, the genome alignment was performed with the program Parsnp v1.2, whose result, locally collinear blocks (LCB) containing the SNPs, were submitted to the program HarvestTools v1.2¹⁶ to extract the alignment.

The phylogeny was performed by the Maximum Likelihood method using the software RaxML v8.0¹⁷, built using the GTR + G model, with supporting nodes calculated using 1000 of bootstrap. The removal of recombinant regions was done using the program ClonalFrameML v1.11¹⁸. The analysis of the population structure was evaluated using hierBAPS algorithm implemented in R language v3.2.3 package¹⁹, using following parameters: a) include singleton SNPs, b) maximum depth = 2 and c) number of populations = 20. Finally, the visualization

of results was performed using the online tool Microreact and accessible following <https://microreact.org/project/9G7UFQLRAwiS4S1J13M69C>²⁰.

Analysis of virulence factors and accessory genome. The BLAST tool was used to search virulence factors using VFDB (Available at <http://www.mgc.ac.cn/VF/>) such as *pefB* (fimbrial regulatory plasmid gene), *shdA* (intestinal colonization adhesin gene), *spvB* (intracellular ADP-ribosylase toxin gene) and *ssaJ* (gene of a major component of Salmonella pathogenicity island 2 type III secretion system), in addition to regions of *S. Enteritidis* ϕ ST64B-like prophage ϕ SE20.

The prediction of coding sequences for each library was performed using the program Prodigal v2.6.3²¹. The homologous genes of the sequences were identified from the genes of the reference genome, with a > 90% identity using BLAST algorithm and > 60% of alignment coverage to the reference. The code used for the annotation is available at <http://github.com/OrsonMM/Blast-score-ratio-for-genomics>. Genomic data have been deposited in GenBank (BioProject ID: PRJNA552561).

Received: 9 March 2020; Accepted: 27 November 2020

Published online: 16 December 2020

References

- Muvhali, M., Smith, A. M., Rakgantsu, A. M. & Keddy, K. H. Investigation of Salmonella Enteritidis outbreaks in South Africa using multi-locus variable-number tandem-repeats analysis, 2013–2015. *BMC Infect Dis.* **17**(1), 661. <https://doi.org/10.1371/journal.pone.0221492> (2017).
- Quesada, A., Reginatto, G. A., Ruiz Español, A., Colantonio, L. D. & Burrone, M. S. Resistencia antimicrobiana de Salmonella spp aislada de alimentos de origen animal para consumo humano. *RevPeruMedExp Salud Publica* **33**(1), 32–44 (2016).
- Campioni, F. *et al.* Changing of the genomic pattern of Salmonella Enteritidis strains isolated in Brazil Over a 48 year-period revealed by whole genome SNP analyses. *Sci. Rep.* **8**(1), 10478. <https://doi.org/10.1038/s41598-018-28844-6> (2018).
- Gantois, I. *et al.* Mechanisms of egg contamination by Salmonella Enteritidis. *FEMS Microbiol.* **33**(4), 718–738 (2009).
- Zhu, J., Li, J. & Chen, J. Survival of Salmonella in home-style mayonnaise and acid solutions as affected by acidulant type and preservatives. *J. Food Prot.* **75**(3), 465–471 (2012).
- Hedican, E. *et al.* Restaurant Salmonella Enteritidis outbreak associated with an asymptomatic infected food worker. *J. Food Prot.* **72**(11), 2332–2336 (2009).
- Le Bacq, F., Louwagie, B. & Verhaegen, J. Salmonella typhimurium and Salmonella enteritidis: changing epidemiology from 1973 until 1992. *Eur. J. Epidemiol.* **10**(4), 367–371 (1994).
- Campos, J. *et al.* Building PulseNet Latin America and Caribbean Salmonella regional database: First conclusions of genetic subtypes of *S. Typhi*, *S. Typhimurium* and *S. Enteritidis* circulating in six countries of the region. *Food Res. Int.* **45**(2), 1030–1036 (2012).
- Dallman, T. *et al.* Phylogenetic structure of European Salmonella Enteritidis outbreak correlates with national and international egg distribution network. *MicrobGenom.* <https://doi.org/10.1099/mgen.0.000070> (2016).
- Deng, X. *et al.* Genomic epidemiology of Salmonella enterica serotype Enteritidis based on population structure of prevalent lineages. *Emerg. Infect. Dis.* **20**(9), 1481–1489 (2014).
- Toro, M. *et al.* Whole-genome sequencing analysis of salmonella entericaserovarEnteritidis isolates in Chile provides insights into possible transmission between gulls, poultry, and humans. *Appl. Environ. Microbiol.* **82**(20), 6223–6232 (2016).
- Grimont, P. & Weill, F. *Antigenic Formulae of the Salmonella serovars*, 9th ed. (Paris: WHO Collaborating Centre for Reference and Research on Salmonella) 1–166 (Institute Pasteur, 2007).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15), 2114–2120 (2014).
- Coil, D., Jospin, G. & Darling, A. E. A5-miseq: an updated pipeline to assemble microbial genomes from IlluminaMiSeq data. *Bioinf. Oxf. Engl.* **31**(4), 587–589 (2015).
- Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**(3), R46. <https://doi.org/10.1186/gb-2014-15-3-r4> (2014).
- Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**(11), 524. <https://doi.org/10.1186/s13059-014-0524-x> (2014).
- Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> (2014).
- Didelot, X. & Falush, D. Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**(3), 1251–1266 (2007).
- Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res.* **3**, 93. <https://doi.org/10.12688/wellcomeopenres.14694.1> (2018).
- Argimón, S. *et al.* Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *MicrobGenom.* **2**(11), e000093. <https://doi.org/10.1099/mgen.0.000093> (2016).
- Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119. <https://doi.org/10.1186/1471-2105-11-119> (2010).
- Moreau, M. R. *et al.* Growth in egg yolk enhances Salmonella Enteritidis colonization and virulence in a mouse model of human colitis. *PLoS ONE* **11**(3), e0150258. <https://doi.org/10.1371/journal.pone.0150258> (2016).
- Arias, I., Zamudio, M. L., Luna, M. A., Valenzuela, W. A. & Cáceres, O. A. Uso de PFGE en la investigación de brotepor Salmonella enteritidis en la localidad de InahuayaRegión Loreto, 2006 Perú. *Boletín Inst. NacSalud.* **14**(9–10), 197–198 (2008).
- Marcelo, M. G. *et al.* Identificación de Salmonella Enteritidis y Salmonella Typhimurium en cuyesmediante la técnica de PCR múltiple. *Rev Investig Vet Perú.* **28**(2), 411–417 (2017).
- Allard, M. W. *et al.* On the Evolutionary history, population genetics and diversity among isolates of Salmonella Enteritidis PFGE pattern JEGX01.0004. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0055254> (2013).
- Zheng, J. *et al.* Genetic diversity and evolution of salmonella entericaSerovarEnteritidis strains with different phage types. *J. Clin. Microbiol.* **52**(5), 1490–1500 (2014).
- Kingsley, R. A., van Amsterdam, K., Kramer, N. & Bäuml, A. J. The *shdA* gene is restricted to serotypes of Salmonella enterica subspecies I and contributes to efficient and prolonged fecal shedding. *Infect Immun.* **68**(5), 2720–2727 (2000).
- Kingsley, R. A., Santos, R. L., Keestra, A. M., Adams, L. G. & Bäuml, A. J. Salmonella enterica serotype TyphimuriumShdA is an outer membrane fibronectin-binding protein that is expressed in the intestine. *MolMicrobiol.* **43**(4), 895–905 (2002).
- Boyen, F. *et al.* The role of SHDA in faecal shedding of Salmonella typhimurium in pigs. *Int ConfEpidemiol Control Biol Chem Phys Hazards Pigs Pork.* 65–67 (2005).

30. Shah, D. H., Zhou, X., Kim, H., Call, D. R. & Guard, J. Transposon mutagenesis of *Salmonella enterica* serovar Enteritidis identifies genes that contribute to invasiveness in human and chicken cells and survival in egg albumen. *Infect Immun.* **80**(12), 4203–4215 (2012).
31. Feasey, N. A. *et al.* Distinct *Salmonella* Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nat Genet.* **48**(10), 1211–1217 (2016).

Acknowledgements

We thank the Earlham Institute and the Global Challenges Research Fund for sequencing facilities, and also the members of the National Reference Laboratory of Enteropathogens/INS, especially Ana Maria Meza Lopez for their technical assistance. This work was supported by the Instituto Nacional de Salud under the project “1000 genomes: Phylogeography and population structure of *Salmonella enterica* in Peru, 1999–2017” with Directorial Resolution No 101-2019-OGITT-OPE/INS.

Author contributions

WQ and RG participated in the conception, hypothesis delineation and study design. CH and MZ conducted isolation, biochemical identification and sequencing of *Salmonella*. JC and OM participated in the analysis, interpretation of data and writing of the article. WQ and RG participated in the critical review of the article.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information is available for this paper at <https://doi.org/10.1038/s41598-020-78808-y>.

Correspondence and requests for materials should be addressed to R.G.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020